

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Кислова Наталья Николаевна

Должность: Проректор по УЧ и качеству образования

Дата подписания: 22.08.2020

Уникальный программный ключ:

52802513f5b14a975b3e9b13008093d5726b159bf6064f865ae65b96a966c035

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

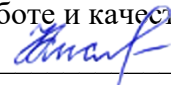
Федеральное государственное бюджетное образовательное учреждение высшего образования

«Самарский государственный социально-педагогический университет»

Кафедра информатики, прикладной математики и методики их преподавания

Утверждаю

Проректор по учебно-методической
работе и качеству образования



Н.Н. Кислова

Макарова Елена Леонидовна

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

для проведения промежуточной аттестации по дисциплине

«Введение в анализ данных государственных органов»

Направление подготовки:

09.03.03 Прикладная информатика

Направленность (профиль):

«Прикладная информатика в государственном и муниципальном управлении»

Квалификация выпускника

бакалавр

Рассмотрено

Протокол № 1 от 25.08.2020

Заседания кафедры информатики, прикладной
математики и методики их преподавания

Одобрено

Начальник Управления образовательных
программ



Н.А. Доманина

Пояснительная записка

Фонд оценочных средств (далее – ФОС) для промежуточной аттестации по дисциплине «Введение в анализ данных государственных органов» разработан в соответствии с ФГОС ВО по направлению подготовки 09.03.03 Прикладная информатика (утвержден приказом Министерства образования и науки Российской Федерации от 19 сентября 2017 г. № 922), основной профессиональной образовательной программой высшего образования по направлению подготовки 09.03.03 «Прикладная информатика» (профиль «Прикладная информатика в государственном и муниципальном управлении»), с учетом требований профессионального стандарта 06.015 «Специалист по информационным системам», утвержденный приказом Министерства труда и социальной защиты Российской Федерации от 18 ноября 2014 г. №896н (зарегистрирован Министерством юстиции Российской Федерации 24 декабря 2014 г., регистрационный №35361), с изменением, внесенным приказом Министерства труда и социальной защиты Российской Федерации от 12 декабря 2016 г. №727н (зарегистрирован Министерством юстиции Российской Федерации 13 января 2017 г., регистрационный №45230).

Цель ФОС для промежуточной аттестации – установление уровня сформированности части компетенций УК-1, ПК-5.

Задачи ФОС для промежуточной аттестации – контроль качества и уровня достижения образовательных результатов по формируемым в соответствии с учебным планом компетенциям:

способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач (УК-1);

способен осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач (ПК-5);

Требование к процедуре оценки:

Помещение: компьютерный класс.

Оборудование: персональные компьютеры с выходом в интернет.

Инструменты: Visual Studio Code/Idle, предустановленные библиотеки NumPy, Pandas, Matplotlib.

Расходные материалы: бумага, ручка.

Доступ к дополнительным справочным материалам: не предусмотрен.

Нормы времени: 90 мин.

Комплект оценочных средств для проведения промежуточной аттестации

Проверяемая компетенция:

Универсальная компетенция УК-1.

Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач.

Проверяемый индикатор:

УК-1.1: анализирует задачу, выделяя этапы ее решения, действия по решению задачи.

Проверяемые образовательные результаты:

Знает: принципы сбора, отбора и обобщения информации, необходимой для анализа открытых данных государственных органов; математические методы анализа открытых данных государственных органов.

Задание 1.

Тип (форма) задания: тест.

Содержание задания:

Вариант 1.

1. К библиотекам анализа данных в Python НЕ относятся ...

а) tkinter

б) Pandas;

в) NumPy;

г) Matplotlib.

2. В качестве переменных для обучения дерева могут быть использованы

а) только количественные признаки;

б) только категориальные признаки;

в) как количественные, так и категориальные признаки;

г) категориальные признаки могут использоваться, только если глубина дерева превышает число переменных, на которых дерево обучалось.

3. Задача классификации сводится к ...

а) нахождения частых зависимостей между объектами или событиями;

б) определения класса объекта по его характеристикам;

в) определение по известным характеристиками объекту значение некоторого его параметра;

г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

4. Задача регрессии сводится к ...

а) нахождения частых зависимостей между объектами или событиями;

- б) определения класса объекта по его характеристиками;
 - в) определение по известным характеристиками объекту значение некоторого его параметра;
 - г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.
5. Задача кластеризации заключается в ...
- а) нахождения частых зависимостей между объектами или событиями;
 - б) определения класса объекта по его характеристиками;
 - в) определение по известным характеристиками объекту значение некоторого его параметра;
 - г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.
6. Целью поиска ассоциативных правил является ...
- а) нахождения частых зависимостей между объектами или событиями;
 - б) определения класса объекта по его характеристиками;
 - в) определение по известным характеристиками объекту значение некоторого его параметра;
 - г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.
7. До предполагаемых моделей относятся такие модели данных:
- а) модели классификации и последовательностей;
 - б) регрессивные, кластеризации, исключений, итоговые и ассоциации;
 - в) классификации, кластеризации, исключений, итоговые и ассоциации;
 - г) модели классификации, последовательностей и исключений.
8. В описательных моделях относятся следующие модели данных:
- а) модели классификации и последовательностей;
 - б) регрессивные, кластеризации, исключений, итоговые и ассоциации;
 - в) классификации, кластеризации, исключений, итоговые и ассоциации;
 - г) модели классификации, последовательностей и исключений.
9. Модели классификации описывают ...
- а) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
 - б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
 - в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
 - г) группы, на которые можно разделить объекты, данные о которых подвергаются анализа.
10. Модели последовательностей описывают ...
- а) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
 - б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
 - в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
 - г) группы, на которые можно разделить объекты, данные о которых подвергаются анализа.
11. Регрессивные модели описывают ...
- а) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
 - б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
 - в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
 - г) группы, на которые можно разделить объекты, данные о которых подвергаются анализа.
12. Какие способы позволяют отобрать 5 первых строк датафрейма `dataf`:
- а) `dataf.head(5); dataf.iloc[0:5]; df.tail(7);`
 - б) `dataf.head(5); dataf.iloc[0:5]; dataf.head(5);`
 - в) `dataf.head(5); dataf.head(5); df.tail(7);`
 - г) `dataf.iloc[0:5]; dataf.head(5); df.tail(7).`
13. Модели исключений описывают ...
- а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;
 - б) ограничения на данные анализируемого массива;
 - в) закономерности между связанными событиями;
 - г) группы, на которые можно разделить объекты, данные о которых подвергаются анализа.
14. Итоговые модели обнаружат ...
- а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;
 - б) ограничения на данные анализируемого массива;
 - в) закономерности между связанными событиями;
 - г) группы, на которые можно разделить объекты, данные о которых подвергаются анализа.

15 Модели ассоциации проявляют ...

а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;

б) ограничения на данные анализируемого массива;

в) закономерности между связанными событиями;

г) группы, на которые можно разделить объекты, данные о которых подвергаются анализа.

16 Виды физической неопределенности данных:

а) неточность измерений значений определенной величины, выполняемых физическими приборами; случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может

стать действительностью)

б) неопределенность значений слов (Многозначность, размытость, непонятность, нечеткость); неоднозначность смысла фраз (Синтаксическая и семантическая);

в) случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью); неопределенность значений слов (многозначность, размытость, неясность, нечеткость);

г) неоднозначность смысла фраз (Синтаксическая и семантическая).

17. Очистка данных — ...

а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д.

б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязку аналитических задач

в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязки аналитического задачи

г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему

18. При обучении дерева решений на обучающей выборке, получена 100% точность классификатора, однако на тестовых классификатор показал точность около 50%. Что может быть причиной такого результата?

а) обучающая выборка слишком большая;

б) возникла проблема переобучения;

в) дерево решений имеет недостаточную глубину;

г) исходные данные некорректны.

19 Консолидация — ...

а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д.

б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязку аналитических задач

в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязки аналитического задачи

г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему

20 Транзакция — ...

а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных

б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов

в) высокоуровневые средства отражения информационной модели и описания структуры данных

г) это установление зависимости дискретной выходной переменной от входных переменных

21 Метаданные — ...

а) некоторый набор операций над базой данных, который рассматривается, как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных

б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов

в) высокоуровневые средства отражения информационной модели и описания структуры данных

г) это установление зависимости дискретной выходной переменной от входных переменных

22 Классификация — ...

а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных;

б) разновидность систем хранения, ориентирована на поддержку процесса

анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость

выполнения аналитических запросов;

- в) высокоуровневые средства отражения информационной модели и описания структуры данных;
- г) это установление зависимости дискретной выходной переменной от входных переменных.

23 Регрессия — ...

- а) это установление зависимости непрерывной выходной переменной от входных переменных;
- б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов;
- в) выявление закономерностей между связанными событиями;
- г) это установление зависимости дискретной выходной переменной от входных переменных.

24 Кластеризация — ...

- а) это установление зависимости непрерывной выходной переменной от входных переменных;
- б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов;
- в) выявление закономерностей между связанными событиями;
- г) это установление зависимости дискретной выходной переменной от входных переменных.

25 Ассоциация — ...

- а) это установление зависимости непрерывной выходной переменной от входных переменных;
- б) эта группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;
- в) выявление закономерностей между связанными событиями;
- г) это установление зависимости дискретной выходной переменной от входных переменных.

26 Машинное обучение — ...

- а) специализированный программный решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных;
- б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов;
- в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат;
- г) подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных.

27 Аналитическая платформа — ...

- а) специализированное программное решение (или набор решений), которое включает в себя все инструменты для извлечения закономерностей из сырых данных;
- б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов;
- в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат;
- г) подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных.

28 Обучающая выборка — ...

- а) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов;
- б) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат;
- в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат;
- г) выявление в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

29 Ошибка обучения — ...

- а) это ошибка, допущенная моделью на учебной множества;
- б) это ошибка, полученная на тестовых примерах, то есть, что вычисляется по тем же формулам, но для тестовой множества;
- в) имена, типы, метки и назначения полей исходной выборки данных;
- г) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат.

30 Ошибка обобщения — ...

- а) это ошибка, допущенная моделью на учебной множества;
- б) это ошибка, полученная на тестовых примерах, то есть, что вычисляется по тем же формулам, но для тестовой множества;
- в) имена, типы, метки и назначения полей исходной выборки данных;
- г) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат.

Правильные ответы к заданию 1

1	А	11	В	21	В
2	В	12	Б	22	Г

3	Б	13	А	23	А
4	В	14	Б	24	Б
5	Г	15	В	25	в
6	А	16	а	26	Г
7	А	17	а	27	А
8	Б	18	Б	28	Б
9	А	19	Г	29	А
10	б	20	А	30	Б

Оценочный лист к заданию 1.

Критерий	Максимальное количество баллов
Вопрос 1-30	1

Проверяемая компетенция:**Универсальная компетенция УК-1.**

Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач.

Проверяемые индикаторы:

УК-1.2: находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи.

УК-1.3: рассматривает различные варианты решения задачи, оценивает их преимущества и риски.

УК-1.4: грамотно, логично, аргументировано формирует собственные суждения и оценки; отличает факты от мнений, интерпретаций, оценок в рассуждениях других участников деятельности.

Проверяемые образовательные результаты:

Умеет: анализировать и систематизировать открытые данные государственных органов с применением информационно-коммуникационных технологий; оценивать эффективность различных математических методов анализа данных государственных органов.

Владет: навыками поиска и практической работы с открытыми данными государственных органов; различными методами решений практических задач дисциплины; методикой оценки результатов решения задач дисциплины.

Способен аргументировать суждения и оценки, опираясь на статистический анализ открытых данных, предоставляемых органами государственного и муниципального управления.

Задание 2.

Тип (форма) задания: практическая задача.

Содержание задания: скачайте с портала «Открытые данные России» Данные Информация об обращениях граждан (сведения о тематике и результатах рассмотрения) <https://data.gov.ru/opendata/7710349494-og>. Подготовьте данные к использованию Средствами языка программирования Python определите наименование вопроса, по которому поступило максимальное количество обращений в Министерство в 4 квартале, и вопросы, по которым поступило минимальное количество обращений.

Оценочный лист к заданию 2.

Указания по оцениванию	Индикатор УК-1	Баллы
Данные скачены; сформирован датасет;	УК1.2	3
Данные очищены, из выборки исключена последняя строка, содержащая суммарное количество обращений;	УК1.3	2
Выведены максимальное количество обращений в Министерство в 4 квартале, и вопросы, по которым поступило минимальное количество обращений.	УК1.4	5
Максимальное число баллов за задание		10

Проверяемая компетенция.**Универсальная компетенция УК-1.**

Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач.

Проверяемый индикатор:

УК-1.5: определяет и оценивает практические последствия возможных вариантов решения задачи.

Проверяемые образовательные результаты:

Способен использовать открытые статистические данные, методы для прогнозирования спроса, продаж программного обеспечения, социально-экономических последствий его внедрения.

Проверяемая компетенция.

Профессиональная компетенция ПК-5.

Способность осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач.

Проверяемые индикаторы:

ПК-5.1: знает принципы, технологии и приемы организации баз данных, проектирования архитектуры информационных систем, нормативный и организационные аспекты управления доступа к данным;

ПК-5.2: умеет проектировать архитектуру ИС различными инструментальными средствами.

Проверяемые образовательные результаты:

Знает принципы, технологии и приемы организации баз данных государственных органов, нормативный и организационные аспекты управления доступа к данным.

Умеет: проектировать компоненты архитектуры ИС, отвечающие за анализ больших данных.

Задание 3

Тип (форма) задания: практическая задача

Содержание задания:

Используя данные со страницы <https://www.kaggle.com/jessemostipak/hotel-booking-demand> (требуется регистрация), определите сколько бронирований было зарегистрировано в каждом из месяцев. Определите есть ли статистически значимая корреляция между длительностью пребывания и временем бронирования ('lead_time'). Обучите дерево для предсказания будет ли отменено бронирование.

Оценочный лист к заданию 3.

Указания по оцениванию	Индикатор УК-1	Баллы
Сформирован датасет; Вычислен новый столбец (длительность пребывания); определен коэффициент корреляции (числовое значение); полученный коэффициент верно интерпретирован;	УК-1.5	5
Данные сгруппированы, определено число бронирований в каждом месяце	ПК-5.1	5
Обучено дерево предсказаний, сделаны выводы.	ПК-5.2	10
Максимальное число баллов за задание		20

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«Самарский государственный социально-педагогический университет»
Методические материалы, определяющие процедуру и критерии оценивания сформированности
компетенций при проведении промежуточной аттестации

Код контролируемой компетенции (индикаторы)	Наименование оценочного средства	Максимальное количество баллов	Всего баллов	Уровень освоения компетенции (в баллах)		
				Пороговый (56-70%)	Продвинутый (71-85%)	Высокий (86-100%)
УК-1.1	Задание 1	30	45	25-32	33-38	39-45
УК-1.2	Задание 2	3				
УК-1.3	Задание 2	2				
УК-1.4	Задание 2	5				
УК-1.5	Задание 3	5				
ПК-5.1	Задание 3	5	15	8-10	11-12	13-15
ПК-5.2	Задание 3	10				

Полученное число баллов выставляется в графу «Промежуточная аттестация» балльно-рейтинговой карты дисциплины.